

Erica Devine
History and New Media
May 6, 2019

**Topic Modeling *Foreign Relations*:
A Distant Reading of U.S Foreign Policy in the Mid-East**

Studying modern diplomatic history in the digital age is both a blessing and a curse. On the one hand, technology has made available digitally a diverse range of sources that would otherwise have been inaccessible, mitigating many of the costs of research, such as travel, housing, and copying. Rich metadata and advanced searching mechanisms have allowed researchers to sift through enormous archives quickly and efficiently. But while the benefits are many, significant challenges have also arisen. The expansion of the foreign policy bureaucracy after World War II led to a significant “archival explosion.”¹ On top of that, the impact of recent historiographical interventions has diversified the types of sources consulted. A scholar studying the history of U.S. foreign policy during the Cold War must examine more traditional documents from the State Department, the executive, Congress, and other government bureaucracies such as the CIA as well as international organization such as the United Nations and the World Bank. The Cultural Turn encouraged the incorporation of less conventional sources, such as travel literature, films, and material culture. The drive towards transnational history encouraged historians to explore sources in other languages and national origins to discuss expressions of U.S. power abroad. The result is that a scholar of foreign relations must read and process a massive amount of literature to fully explore their chosen topic.

¹ David Allen and Matthew Connelly in David Allen, “Diplomatic History After the Big Bang: Using Computational Methods to Explore the Infinite Archive,” in *Explaining the History of American Foreign Relations*, ed. Matthew Connelly, 3rd ed. (Cambridge: Cambridge University Press, 2016), 78.

In the archives themselves, rich metadata has allowed scholars to quickly search large archives and even to search the texts themselves if they are digitized. But searching mechanisms are not always effective. If the documents lack metadata or the term you are searching is more abstract, then a searching mechanism will not be effective. This system also has an element of archival bias: metadata is created by archivists, who determine what is important in a given document. As a result, the expansion of the archive and the difficulties with searching mechanisms have made it difficult for scholars to locate, close read, and analyze all the documents on their topic.

Recently, researchers in other fields have been exploring ways to use digital tools to mine the vast archive. One such method is topic modeling. This tool allows scholars to generate an index of a vast corpus of digital texts. It locates a set of topics and themes within the texts. From the results of a topic model, scholars can locate which individual documents contain that topic and can show how those topics change over time. Topic modeling is not an analytical tool in and of itself, but when used as an index tool and in conjunction with traditional scholarship, it can open up new avenues of inquiry. Diplomatic historians have just begun to explore computational methods as a compliment to traditional scholarship, but have not yet attempted to explore in depth how topic modeling can be used to deal with the massive expansion of the archives.

This article will explore topic modeling as a methodological tool in the study of American foreign relations with the Middle East during the first half of the 1950s. The first section will review the scholarship on macroanalysis in general and topic modeling in particular: what it is and how it works, how it has been used as a tool of historical research, and how it has been discussed by diplomatic historians. Next, I will elaborate on my corpus and detail my

methodology. Finally, I will present my results, preliminary interpretations, challenges, and new questions that arose from my analysis.

Literature Review

Scholars across a number of disciplines have turned to macroanalysis to mine large swaths of text. Macroanalysis or “distant reading” is a technique that uses digital technology to approach large-scale research in innovative ways. Franco Moretti and Matthew Jockers have applied this concept to their work in literary studies. Moretti, an Italian literary scholar, argues that a close reading of texts depends on a small corpus of documents, which limits the scholars’ ability to address themes and tropes on a larger scale. He proposes the scholars turn to computational textual analysis to aggregate and analyze large amounts of data.² In his book *Macroanalysis*, Jockers takes up this task. He investigates different digital methods to shift “from looking at the individual occurrences of a feature in context to looking at the trends and patterns of that feature aggregated over an entire corpus.”³ His study explores the ways in which literary scholars can use computational analysis to scan metadata, map nationality, and discover style, theme, and influence across a corpus of thousands of books. He concludes, like Moretti, that macroanalysis can unearth new forms of digital evidence and uncover new questions and avenues for exploration.⁴

Topic modeling is one method of distant reading that Jockers investigates. Latent Dirichlet Allocation (LDA) topic modeling is a statistical model that identifies a topic or topics that appear in a single document. When applied to a large corpus of texts, it generates clusters of

² Franco Moretti, *Distant Reading* (New York: Verso, 2013), 48–49 Moretti is discussing the necessity of distant reading to study world literature, which has a corpus too large to perform a close reading of even half of the texts in this corpus.

³ Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History* (Urbana and Chicago: University of Illinois Press, 2013), 24.

⁴ *Ibid.*, 10.

words that share a topic or theme and determines which documents share themes based on which words are likely to “co-occur in them.”⁵ The number of topics generated each time is arbitrary and determined by the scholar. The standard is twenty but larger corpus should warrant a larger number of topics.⁶ Most significantly, topic models generate topics without human input: you do not need to have pre-existing schema to run the program. This feature minimizes researcher bias by making pre-existing categories unnecessary and it can potentially reveal themes and trends that the researcher may not have been aware of. Once the topics are generated, however, the topics need to be interpreted by the scholar and given a label. While it is not necessary for the scholar to have read all the sources in the corpus, they need to be familiar with the subject matter to be able to interpret the results.

Historian Sharon Block, along with computer scientists David Newman at the University of California Irvine, uses topic modeling to explore the entire eighteenth century corpus of the *Pennsylvania Gazette*. Her corpus contained approximately eighty-two thousand articles and advertisements from 1728 to 1800. Using her results, she walks readers through how digital analysis can be used as a tool alongside “the historians’ input and analysis.”⁷ Her research reveals new information about word frequency and word use. For example, she notices that the word “slavery” appears most often with words relating to revolutionary ideals and government forms, which she suggests indicates that the readers of this newspaper were more likely to see discussions of slavery in relation to “the rhetoric of political enslavement than in relation to the actual enslavement of Africans.”⁸ More recently, Cameron Blevins explores the benefits of topic

⁵ Sharon Block, “Doing More with Digitization: An Introduction to Topic Modeling of Early American Sources,” *Common-Place The Interactive Journal of Early American Life, Inc.* 6, no. 2 (January 2006), <http://www.common-place-archives.org/vol-06/no-02/tales/>.

⁶ Jockers, *Macroanalysis*, 123.

⁷ Block, “Doing More with Digitization”

⁸ *Ibid.*

modeling. In his topic model of Martha Ballard's diary, a corpus of nearly 10,000 diary entries, he similarly identifies thematic trends and unseen patterns of word usage, which reveal information about the texts that could not be garnered from a close reading.⁹ For example, in the topic Blevins labeled death, words like “informd,” “hear,” and “heard” appeared alongside words like “morn” and “death,” indicating that death is often encountered through face-to-face interactions in the form of news. As both Block and Blevins demonstrate, topic modeling has the potential to reveal patterns in a massive data set that would otherwise have been lost.¹⁰

Amanda Regan, a doctoral candidate at George Mason University, is currently utilizing topic modeling in her dissertation on how ideas about physical fitness for women have changed. She ran a topic model of the fitness magazine *Mind and Body* from 1894 to 1936. She discovers shifts in the standards of physical education and a change in the relationship between hygiene, physical fitness, and disease over time.¹¹ Her advice to historians is particularly important. As historians in the “digital present,” we must be critical of our methodologies when working with online sources and transparent in their use, as she attempts to do in her article. She concludes that “research methodologies in the digital present must maintain pace with the accelerated volume of digitized material, and scholars must adapt their research strategies to fit this digital environment.”¹²

The consensus among scholars is that topic modeling is an innovative methodological tool that can generate new fields of inquiry. But topic modeling is not without flaws. Jockers discusses the drawbacks of this tool. After examining a model of 3,346 books, he noticed that

⁹ Cameron Blevins, “Topic Modeling Martha Ballard’s Diary,” April 1, 2010, <http://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/>.

¹⁰ Ibid.

¹¹ Amanda Regan, “Mining Mind and Body: Approaches and Considerations for Using Topic Modeling to Identify Discourses in Digitized Publications,” *Journal of Sport History* 44, no. 2 (2017): 174.

¹² Ibid.

most of the topics developed by the algorithm were easily interpretable. But some were either too general or incoherent. These types of topics, what he calls “black boxes,” are often the product of unsupervised models. The machine generates topics with no concern for the interpretability of the information so it is very likely that some will be incomprehensible, especially if stop words are not removed. He argues, however, that it is completely legitimate to ignore “black boxes” or general topics without compromising the integrity of the model.¹³

In the field of diplomatic history, computational analysis has generated some attention. In the third edition of *Explaining the History of American Foreign Policy*, David Allen and Matthew Connelly explore the role of computational methods in navigating a dramatically increasing foreign policy archive. These methods, according to the authors, “may offer the only hope of creating order from the chaos and producing anything like a proper finding aid” in archives that are increasingly underfunded and swamped with new digital material.¹⁴ Diplomatic history lends itself well to computational analysis. The state department publications provide machine-readable text with rich metadata to the public. Allen and Connelly provide the readers with a brief example of a topic model of several *Foreign Relations of the United States* (FRUS) volumes related to the Kennedy and Johnson administrations’ Vietnam policies. Topic modeling, they conclude, reveals the “hidden intellectual structures” in a corpus and can demonstrate how those structures change over time.¹⁵ Unfortunately, the authors do not analyze the results in depth but rather use it to exemplify the possibilities behind this method.

Methodology

Here, I will take up Allen and Connelly’s call to incorporate computational analysis into the study of American foreign relations. For my initial foray into topic modeling, I started with

¹³ Jockers, *Macroanalysis*, 128–30.

¹⁴ Allen, “Diplomatic History After the Big Bang,” 83.

¹⁵ *Ibid.*, 90.

Foreign Relations of the United States (FRUS), the official documentary historical record of major U.S. foreign policy decisions. This publication, produced by the Office of the Historian, contains documents from Presidential libraries, the Department of State and Defense, the National Security Council, the Central Intelligence Agency, the Agency for International Development, as well as private papers from policymakers.¹⁶ Each volume in the series is organized chronologically by Presidential administration, geographically and topically. I looked at three volumes of FRUS covering the period 1951-1954, focusing on the Near East and Iran. These volumes consisted of 1,922 documents covering policy related to Egypt, Iran, Syria, Lebanon, Jordan, Iraq, and Saudi Arabia.¹⁷

I used MALLET, or MACHine Learning for Language Toolkit, to run the topic model. It was created by Andrew McCallum at the University of Massachusetts, Amherst and is defined as “a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text.”¹⁸ To run the program, I first placed each document in an individual word document and labeled it with the document number from FRUS (the reason I mention this will become clear when I discuss my results). The text included the date and title but excluded all footnotes because they often contain references to other documents or the editor’s own comments. Once the data was imported into MALLET, I chose to run a forty-topic analysis.

¹⁶ Office of the Historian, Bureau of Public Affairs, “About Us,” Office of the Historian, 2019, <https://history.state.gov/about>.

¹⁷ *Foreign Relations of the United States, 1952–1954, Iran, 1951–1954, Volume X* (Washington, D.C: United States Government Printing Office, 1989); *Foreign Relations of the United States, 1952–1954, The Near and Middle East, Volume IX, Part 1* (Washington, D.C: United States Government Printing Office, 1986); *Foreign Relations of the United States, 1952–1954, The Near and Middle East, Volume IX, Part 2* (Washington, D.C: United States Government Printing Office, 1986). Three documents are also related to Kuwait and Yemen.

¹⁸ McCallum, Andrew Kachites. “MALLET: A Machine Learning for Language Toolkit.” <http://mallet.cs.umass.edu>. 2002.

```
Ericas-MacBook-Air-2:~ Erica$ cd /Users/Erica/Desktop/mallet
Ericas-MacBook-Air-2:mallet Erica$ ./bin/mallet import-dir --input sample-data/w
eb/frus --output tutorial.mallet --keep-sequence --remove-stopwords
Labels =
  sample-data/web/frus
Ericas-MacBook-Air-2:mallet Erica$ bin/mallet train-topics --input tutorial.mal
let --num-topics 40 --output-state topic-state.gz --output-topic-keys tutorial
_keys.txt --output-doc-topics tutorial_composition.txt
Mallet LDA: 40 topics, 6 topic bits, 111111 topic mask
Data loaded.
max tokens: 4574
total tokens: 579426
```

FIGURE 1: This is a screenshot of the code I used to run the topic model.

As I mentioned above, this number is arbitrary. However, after running four different models to generate five, twenty, forty, and one hundred topics, I settled on forty. Generating only five topics was not enough to do a proper analysis. On the other hand, one hundred topics produced so many topics, many of which were too specific and often fell into similar categories, that an analysis was difficult. My advice is to generate between thirty and fifty topics. Even though some of the results may be too general or uninterpretable, you will still have enough data to work with.

Results, Interpretations, and Challenges

A completed topic model produces two documents. Document 1 is the list of topics with the percentage of the corpus that is associated with that topic. Document 2 lists each document in the corpus and the percentage of that document associated with each topic.

Document 1

| | | |
|----|--------|---|
| 0 | 12.88% | military aid assistance program defense equipment arms |
| 1 | 2.69% | general middle command east nato group turkey greece |
| 2 | 5.76% | secretary dulles president world people countries east st |
| 3 | 7.46% | jordan israel border mac israeli tel aviv incidents arab an |
| 4 | 2.10% | oil companies petroleum foreign american report suit ca |
| 5 | 7.63% | iran iranian oil shah aid government british henderson ki |
| 6 | 11.17% | defense east middle states arab area military organizatic |
| 7 | 3.01% | egypt sudan sudanese king egyptians eden title egyptian |
| 8 | 12.50% | egypt egyptian egyptians base british cairo caffery agree |
| 9 | 45.15% | situation page time long hope attitude relations good ac |
| 10 | 4.34% | jordan syria syrian israel water development project plar |
| 11 | 36.83% | agreement agreed draft points negotiations paragraph p |
| 12 | 11.79% | secretary president state memorandum states united ge |
| 13 | 44.52% | embassy department secretary state washington p.m sei |
| 14 | 4.49% | naguib army rcc officers tudeh general regime coup mah |
| 15 | 21.39% | british foreign office eden negotiations london ambassac |
| 16 | 4.77% | council resolution israel security armistice bennike zone |
| 17 | 7.42% | oil company price gulf production crude companies kuwi |

Document 2

| | | | | |
|--------------------|-----|----|----|----|
| 1\1.txt | 5% | 0% | 0% | 0% |
| 2\100%20copy.txt | 0% | 0% | 0% | 0% |
| 3\10.txt | 0% | 0% | 0% | 5% |
| 4\100%20copy.txt | 0% | 0% | 0% | 0% |
| 5\100.txt | 0% | 0% | 0% | 0% |
| 6\1000%20copy.txt | 29% | 0% | 0% | 0% |
| 7\1001%20copy.txt | 4% | 0% | 0% | 0% |
| 8\1002%20copy.txt | 0% | 0% | 0% | 0% |
| 9\1003%20copy.txt | 33% | 0% | 0% | 0% |
| 10\1004%20copy.txt | 12% | 0% | 0% | 0% |
| 11\1005%20copy.txt | 1% | 0% | 0% | 2% |
| 12\1006%20copy.txt | 2% | 0% | 0% | 0% |
| 13\1007%20copy.txt | 0% | 0% | 0% | 0% |
| 14\1008%20copy.txt | 0% | 0% | 0% | 0% |

I excluded some topics from my analysis right away, despite their high percentage rate, because they were too general. For example, topics labeled Embassies or Communications were passed over, despite their frequency in the corpus. Other topics were black boxes, like “tilocblob dsdb ocblob bud.” Taking Jockers’ advice, I passed over these topics in favor of focusing on “the most interpretable.”¹⁹ Next, I will walk you through my results, preliminary interpretations, the challenges I faced, and some new avenues for exploration.

¹⁹ Jockers, *Macroanalysis*, 129.

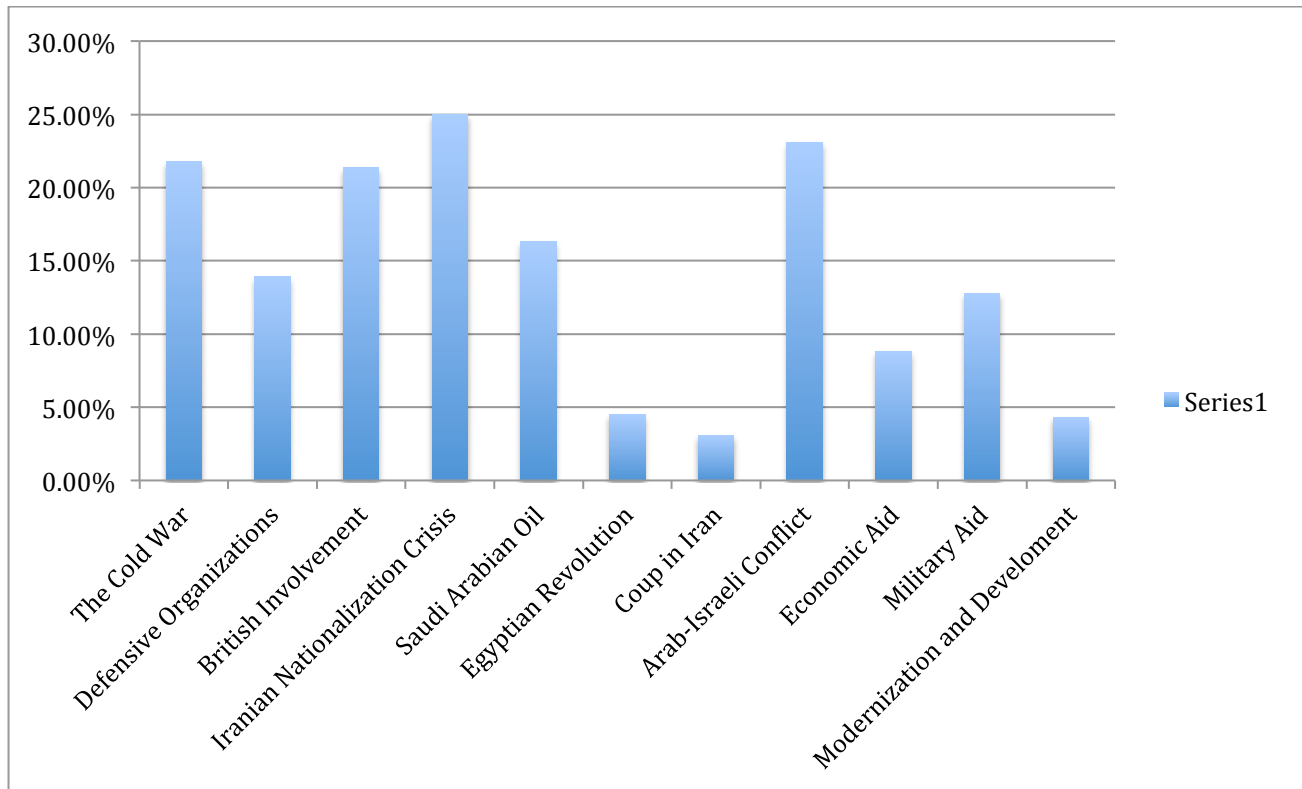


Figure 2: This chart marks the percentage of the corpus associated with each topic. Some topics, like oil, the Arab-Israeli Conflict, and Defensive Organizations are the combination of several topics associated with similar themes.

One of the topics with the highest frequency I labeled “The Cold War.” The presence and frequency of this topic was not surprising considering the historical context. It demonstrates that Cold War concerns were paramount in diplomatic considerations between 1951 and 1954. The co-occurrence of words like “economic,” “security,” “military,” and “strength” intertwined with “free world” point to U.S. strategy in the early Cold War. The U.S. policy of containment, solidified in the Truman Doctrine in 1947, emphasized maintaining access to oil, providing economic and military aid to pro-American regimes, and stifling home-grown nationalist movements, especially ideologically communist or socialist, that threatened American interests.²⁰ Along these lines, the topic that I labeled “Defensive Organizations” speaks to the prevalence of this rhetoric in foreign policy planning. As the U.S. recognized the waning power of Britain in

²⁰ See Odd Arne Westad, *The Cold War: A World History* (New York: Basic Books, 2017).

the region, efforts were made for the U.S. to replace them. To secure petroleum-producing regions, prevent the spread of communism, and establish regional stability, the Policy Planning Staff at the State Department proposed the creation of a Middle Eastern Defense Organization (MEDO) similar to NATO. Most of the documents in this corpus focus on U.S. attempts to convince Egypt to join MEDO, though their attempts were unsuccessful.²¹ MEDO failed, but forming regional defense organizations continued to be a primary goal of U.S. foreign policy, resulting in the Baghdad Pact in 1955 and Central Treaty Organization (CENTO) in 1958.

| | | |
|--------------------------------|--|-------|
| The Cold War | economic soviet political security area forces western military support world policy free measures strength influence communist internal ussr west power | 21.8% |
| Defensive Organizations | defense east middle states arab area military organization arrangements powers medo u.s regional planning participation turkey western forces arrangement conference | 11.1% |
| | pakistan turkey turks india turkish turk military nehru defense afghanistan aid pakistanis indian ankara countries pakistani karachi arrangement pact pak | 2.8% |

The model also produced themes relating to “Economic Aid,” “Military Aid,” and “Modernization and Development.” The topics speak to the implementation of modernization theory in the Middle East. Granting either economic or military aid was a way for the U.S. to secure allies in the region and bolster pro-American regimes. The two forms of aid went hand in hand with the acceptance of modernization theory. Modernization refers to ideas about how best to transform society from pre-modern or traditional to modern. Modernization played a large role in interactions between the U.S. and the Third World in the postwar world. Proponents of this theory believed that countries that were modernized would be created in America’s image and therefore more stable and less likely to turn to communism. Similarly, military modernization was intertwined with economic aid, as both an enticement to remain aligned to the West and to

²¹ See documents 995, 1010, 1029, 1048, 1050, 1056, in *FRUS, 1952–1954, The Near and Middle East, Volume IX, Part 2*.

build up the military might of Western allies. For example, in Egypt, severe economic issues related to a dependence on cotton as a cash crop prompted U.S. development assistance. American Ambassador to Egypt, Jefferson Caffrey, wrote to the State Department requesting \$100 million for “economic development and industrialization” projects such as the construction of a new Aswan dam, “Hydroelectric installations,” establishing a steel and iron industry, and building infrastructure.²²

| | | |
|--------------------------------------|---|-------|
| Economic Aid | million loan aid economic bank funds program assistance financial year development projects foreign exchange fiscal grant ibrd debt technical request | 8.8% |
| Military Aid | military aid assistance program defense equipment arms egypt grant department economic mutual security training state list act items supply million | 12.8% |
| Modernization And Development | jordan syria syrian israel water development project plan johnston waters valley river yarmuk land shishakli zone economic damascus dam moose | 4.3% |

The topic model also drew my attention to facets of those topics that I would have overlooked during a close reading of the texts. For example, the concern for British influence throughout the corpus indicates that Britain was still very much a factor in Middle Eastern affairs during these years. While the U.S. and Britain were not competing militarily, there were economic tensions across the region. In Egypt, the U.S. was negotiating with the new Revolutionary Egyptian government over the British presence in the Suez Canal Zone. In the Gulf, the British and the U.S. clashed over the Buraimi dispute, with the U.S. backing their Saudi allies at the expense of the British. Finally, the U.S. worked to maintain British access to petroleum in Iran and Iraq. Scholars have differing opinions on the role of Britain in the Middle East during this time period: some argue they took a subservient role, others that they were active players. Some, like Mary Ann Heiss argue that the British role declined after the Iranian oil crisis in 1953, others contend that the decline occurred during the Suez Crisis in 1956. Stephen Galpern argues that it was not until the depreciation of sterling in 1967 that the British

²² The Ambassador in Egypt (Caffery) to the Department of State, December 11, 1952, in *FRUS, 1952–1954, The Near and Middle East, Volume IX, Part 2*, documents 1046.

Empire in the Middle East ended. The frequency with which the British appeared in this corpus supports the scholarship that contends that Britain was still a major force in the region. For future projects, I would be interested to see how the British presence changed over time and when.²³

| | | |
|----------------------------|--|-------|
| British Involvement | british foreign office eden negotiations london ambassador proposal kingdom point position united embassy department talks meeting government proposals agreed | 21.4% |
|----------------------------|--|-------|

The topic that appeared most frequently was oil. Nine of the forty topics generated were related in some way to oil: Saudi oil, Iranian oil, the oil industry, oil agreements, oil transportation, etc. In the topic Iranian Oil, words like “British” and “settlement” indicate American concern over Prime Minister Mossadegh’s nationalization of the Iranian oil supply, at the expense of the Anglo-Iranian Oil Company (AIOC). The inclusion of words like “world” and “future” are particularly telling. It speaks to the way the U.S. viewed maintaining access to Iranian oil. U.S. foreign policy officials consistently framed negotiations over nationalization in terms of what was best for the future of Iran. In a letter from Truman to Mossadegh in 1951, Truman writes that the oil dispute was “so full of dangers to the welfare of your own country” and advises Mossadegh to consider the “future well-being and prosperity” of the country.²⁴ The first topic in the category also highlights the symbiotic relationship between private oil companies and the U.S. government which scholars have been exploring for some time: the U.S. sponsored a coup which overthrew Mossadegh, which allowed the five major American oil companies to gain a portion of the Iranian concession. The “cartel” was challenged domestically

²³ See Mary Ann Heiss, *Empire and Nationhood: The United States, Great Britain, and Iranian Oil, 1950-1954* (New York: Columbia University Press, 1997); Steven G. Galpern, *Money, Oil, and Empire in the Middle East: Sterling and Postwar Imperialism, 1944-1971* (Cambridge: Cambridge University Press, 2009).

²⁴ President Truman to Prime Minister Mosadeq, July 8, 1951, in *FRUS, 1952-1954, Iran, 1951-1954, Volume X*, document 37.

in an anti-trust case, but the U.S. government, recognizing the Cold War strategic need to maintain access to foreign oil, allowed it to remain.²⁵

| | | |
|---|---|------|
| Iranian Oil Nationalization Crisis | oil companies petroleum foreign american report suit case national cartel commission international middle world east civil anti-trust operations action interests | 2.1% |
| | iran iranian oil shah aid government british henderson kind army make public give problem present future world financial settlement country | 7.6% |
| | oil compensation company iran mosadeq arbitration aioc court iranian henderson international million contract proposals dmpa terms tehran law claims payment | 4.6% |
| | shah mosadeq ala majlis zahedi iran tehran country court tudeh kashani henderson senate front elections oil opposition support successor political | 3.1% |
| | iran oil mosadeq iranian bank harriman iranians nationalization tehran aioc irans president company govt brit proposal industry top technicians secretary | 4.1% |
| | consortium iran aioc iranian companies hoover government iranians tehran participation hmg group london oil negotiations management american matter representatives distribution | 3.5% |

Conversely, discussions about Saudi oil focuses on three areas of tension in the region. The first was over the transportation of Saudi oil by sea. U.S. efforts during this time were applied towards establishing an American monopoly over the production and transportation of Saudi oil and thus eliminating foreign middlemen. In 1954, the National Security Council declared that “The United States should take all appropriate measures to bring about the cancellation of the agreement between the Saudi Arabian Government and Onassis [a Greek shipping company] for the transport of Saudi Arabian-produced oil and, in any case, to make the agreement ineffective.”²⁶ The U.S. was also concerned about regulating petroleum prices on the world market and supporting Saudi Arabia in their conflict against the other Gulf States and the British over the Buraimi oasis. Finally, a major point of contention between the U.S. and the other Arab states was the construction of a trans-Arabian pipeline (TAPLINE) across non-oil producing states, such as Jordan and Syria.²⁷

²⁵ See David S Painter, *Oil and the American Century: The Political Economy of U.S. Foreign Oil Policy, 1941-1954* (Baltimore: Johns Hopkins University Press, 1986); Irvine H. Anderson, *Aramco, the United States, and Saudi Arabia: A Study of the Dynamics of Foreign Oil Policy, 1933-1950* (Princeton: Princeton University Press, 1987).

²⁶ Statement of Policy by the National Security Council, July 23, 1954, in *FRUS, 1952-1954, The Near and Middle East, Volume IX, Part 1*, document 219.

²⁷ See Painter, *Oil and the American Century*; Nathan. J. Citino, *From Arab Nationalism to OPEC: Eisenhower, King Saud, and the Making of U.S.-Saudi Relations* (Bloomington: Indiana University Press, 2002).

| | | |
|--------------------------|--|------|
| Saudi Arabian Oil | aramco saudi onassis agreement oil arabia sag department company arabian concession companies duce area king disputed government davies tankers operations | 2.8% |
| | oil company price gulf production crude companies kuwait countries basis prices products petroleum industry pay operations percent years income concession | 7.4% |
| | companies oil ipc middle east government company aramco payments beeley transit countries pipeline tapline iraq governments negotiations saudi eakens u.s | 2.2% |
| | saudi arabia buraimi king british prince saudis area arbitration sag faisal dispute jidda saud arabian ambassador hare dhahran ibn boundary | 3.9% |

The topic model identified two topics related to coups that occurred in the Middle East during this time period. The first took place in Egypt in 1952. A group of army officers led by Mohammed Naguib and Gamal Abdel Nasser overthrew the British-backed monarch, King Farouk, and established a republic predicated on Arab nationalism and anti-imperialism.²⁸ The Iranian coup was sponsored and funded by the CIA to overthrow the democratically elected Prime Minister Mohammed Mossadegh in Iran after his nationalization of the Iranian oil industry.²⁹ The presence of “tudeh” in the topic labeled “Egyptian Revolution” was surprising at first. The Tudeh Party was an Iranian Marxist-Leninist party that joined Mossadegh’s coalition and promoted nationalization of Iranian oil. Most scholars content that the Tudeh’s presence in the coalition was a significant factor in the U.S. decision to overthrow Mossadegh, as they feared he was a communist or a communist supporter.³⁰ Seeing the word “tudeh” associated with Egypt peaked my curiosity. Was the U.S. government concerned that the Revolutionary Command Council in Egypt was similarly under the influence of communist forces, like in Iran? After exploring the documents themselves, I found a more mundane reason for the co-occurrence. Discussions of the “tudeh” and Egypt co-occur in National Intelligence Estimates and in Memorandum of the National Security Council meetings. For example, in an intelligence

²⁸ See Salim Yaqub, *Containing Arab Nationalism: The Eisenhower Doctrine and the Middle East* (Chapel Hill: University of North Carolina Press, 2004).

²⁹ See Chapter Six of Nikki R Keddie and Yann Richard, *Modern Iran: Roots and Results of Revolution* (New Haven: Yale University Press, 2006).

³⁰ See Heiss, *Empire and Nationhood*; Painter, *Oil and the American Century*.

estimate on current security threats in the region, after an assessment of the Egyptian situation, the report states, “Iran presents a more pressing problem than that existing in the other states of the area, owing in part to the proximity of the Soviet Union and the strength of the Tudeh Party, and in part to the more immediate danger of social, political, and fiscal breakdown.”³¹ These terms co-occurred so often because they were often discussed in the same document.

| | | |
|----------------------------|---|------|
| Egyptian Revolution | naguib army rcc officers tudeh general regime coup maher party august elements control embassy forces staff ali caffery july cabinet | 4.5% |
| Coup in Iran | shah mosadeq ala majlis zahedi iran tehran country court tudeh kashani henderson senate front elections oil opposition support successor political | 3.1% |

The most striking aspect of this topic model is the absence of religion. Recent scholarship has focused on the role of religion in U.S. foreign policy in the early Cold War.³² In particular, these scholars have explored the religious rhetoric adopted by figures such as President Eisenhower and John Foster Dulles in relations with Arab leaders, indicating a strong belief that appealing to the Muslim world would combat “godless” Soviet inroads in the region. However, none of the topics generated by the model focused on words relating to religion. I took a closer look at the corpus itself to see if religious language was completely absent, or if words did not co-occur together enough to register. I conducted a simple keyword search for the words “religion,” “Islam,” “Moslem,” and “Christian” across all three volumes. What I found confirmed that religion was significant in a number of ways. It was used as a way of referencing states in the Middle East: the terms “Moslem world” or “Moslem states” are used

³¹ National Intelligence Estimate, January 15, 1953, in *FRUS, 1952–1954, The Near and Middle East, Volume IX, Part 1*, document 114.

³² See Andrew Preston, *Sword of the Spirit, Shield of Faith: Religion in American War and Diplomacy* (New York: Knopf Doubleday Publishing Group, 2012); William Inboden, *Religion and American Foreign Policy, 1945-1960: The Soul of Containment* (Cambridge University Press, 2010).

interchangeably with “Arab” and “north Africa.” In discussions of the Arab-Israeli conflict, references were made to the shared “Judeo-Christian civilization which is America’s heritage.”³³

Throughout these documents, Islam was considered both an ally and a threat depending on the context. In a memorandum from Hoskins to Byroade, he wrote:

The tides of neutralism and nationalism, with which communism has successfully allied itself to an increasing degree, continue to rise. One impressive development since my trip last year is the surprising increase in the number of acknowledged communists in the Middle East and their growing influence, despite the Moslem religion which many, including King Ibn Saud, felt would form an effective barrier.³⁴

One such group was al-Ikhwān al-Muslimīn, or the Muslim Brotherhood, which spread first across Egypt then into other states through the construction of mosques, schools, and educational and welfare services, establishing an organizational network that mobilized large segments of the population in mass riots against the government. The Egyptian branch in the 1940s was believed to have over two million members.³⁵ In five documents concerning the Muslim Brotherhood, they were associated with “commies” and other opposition groups, such as the Wafd Party.³⁶ In Iran, the “fanatical Moslem leader” Kashani was watched closely when he marshaled his “fanatical followers” against Mossadegh in the Majlis and eventually joined the CIA payroll during the coup in August 1953.³⁷ Religion was, in fact, part of the dialogue during these years, despite the results of the topic model.

³³ Memorandum of Conversation, by the Director of the Office of Near Eastern Affairs (Hart), October 8, 1953, *FRUS, 1952–1954, The Near and Middle East, Volume IX, Part 1*, document 683.

³⁴ Memorandum by the Acting Regional Planning Adviser, Bureau of Near Eastern, South Asian, and African Affairs (Hoskins), to the Assistant Secretary of State for Near Eastern, South Asian, and African Affairs (Byroade), July 25, 1952, in *FRUS, 1952–1954, The Near and Middle East, Volume IX, Part 1*, document 81.

³⁵ For more information on the Muslim Brotherhood, see Lefèvre, *Ashes of Hama*; Elizabeth F. Thompson, “Hasan Al-Banna of Egypt: The Muslim Brotherhood’s Pursuit of Islamic Justice,” in *Justice Interrupted: The Struggle for Constitutional Government in the Middle East* (Cambridge: Harvard University Press, 2013), 150–76; Richard Paul Mitchell, *The Society of the Muslim Brothers* (Oxford: Oxford University Press, 1993).

³⁶ See documents 997, 1001, 1020, 1033, 1218, in *FRUS, 1952–1954, The Near and Middle East, Volume IX, Part 2*.

³⁷ Memorandum Prepared in the Office of National Estimates, Central Intelligence Agency, for the President, March 1, 1953, in *FRUS, 1952–1954, Iran, 1951–1954, Volume X*, document 310.

Conclusion

This research has been primarily exploratory in nature. My goal was to investigate how topic modeling can be used as a methodological tool in the field of diplomatic history. One of the ways that this tool has been particularly useful is in identifying important themes throughout the corpus. The topic model was spot on when grouping words related to the most prominent themes in U.S. foreign policy during these years: the prominence of Cold War concerns, the strategic importance of oil, the significant role of modernization and development theory, tied closely with military and economic aid, and the dominant role of the British in the region. In other areas, it came up short: religion played a much larger role than the topic model indicated. This highlights the importance of combining this type of distant reading with a close reading of the texts. Topic modeling cannot replace more traditional historical methods: as Sharon Block writes, “Archives will still be visited and documents will still be read. But this new technique may allow scholars to use digital archives not just to *access* increasing numbers of documents but to *analyze* those documents in entirely new ways.”³⁸

One way of mapping topic models that was not touched on in this article is showing how topics change over time. To do this type of analysis, each topic must be labeled by date, which I did not do initially. Being able to map topics over time allows scholars to trace the rise and fall of particular themes.³⁹ While this mechanism would be less effective with such as chronologically short corpus, the prospect of mapping a corpus that covers multiple decades has the potential to reveal trends that would be overlooked with just a close reading.

Overall, this project has demonstrated several things. First, topic modeling is very effective at generating prominent themes the run through a large corpus. As I demonstrated

³⁸ Block, “Doing More with Digitization.”

³⁹ For examples of this, see Blevins, “Topic Modeling Martha Ballard’s Diary”; Allen, “Diplomatic History After the Big Bang.”

above, the topics confirmed things I already knew about the time period. Secondly, the combination of topic model and close reading allowed me to explore questions that arose as a result of the model. For example, the presence of “tudeh” and the absence of religion prompted me to investigate further. While these new directions confirmed, more than challenged, an already held view, it still demonstrates how a topic model can get you to think about a subject in a new way. Lastly, this research has encouraged me to pursue this project on a larger scale. The application of topic modeling across the entire span of FRUS documents during the Cold War could be very revealing, especially in demonstrating how topics have changed over time.

Bibliography

- Allen, David. "Diplomatic History After the Big Bang: Using Computational Methods to Explore the Infinite Archive." In *Explaining the History of American Foreign Relations*, edited by Matthew Connelly, 3rd ed. Cambridge: Cambridge University Press, 2016.
- Amanda Regan. "Mining Mind and Body: Approaches and Considerations for Using Topic Modeling to Identify Discourses in Digitized Publications." *Journal of Sport History* 44, no. 2 (2017): 160–77.
- Anderson, Irvine H. *Aramco, the United States, and Saudi Arabia: A Study of the Dynamics of Foreign Oil Policy, 1933-1950*. Princeton: Princeton University Press, 1987.
- Blevins, Cameron. "Topic Modeling Martha Ballard's Diary," April 1, 2010. <http://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/>.
- Block, Sharon. "Doing More with Digitization: An Introduction to Topic Modeling of Early American Sources." *Common-Place The Interactive Journal of Early American Life, Inc* 6, no. 2 (January 2006). <http://www.common-place-archives.org/vol-06/no-02/tales/>.
- Citino, Nathan. J. *Envisioning the Arab Future: Modernization in US-Arab Relations, 1945-1967*. Cambridge: Cambridge University Press, 2017.
- . *From Arab Nationalism to OPEC: Eisenhower, King Saud, and the Making of U.S.-Saudi Relations*. Bloomington: Indiana University Press, 2002.
- Foreign Relations of the United States, 1952–1954, Iran, 1951–1954, Volume X*. Washington, D.C: United States Government Printing Office, 1989.
- Foreign Relations of the United States, 1952–1954, The Near and Middle East, Volume IX, Part 1*. Washington, D.C: United States Government Printing Office, 1986.
- Foreign Relations of the United States, 1952–1954, The Near and Middle East, Volume IX, Part 2*. Washington, D.C: United States Government Printing Office, 1986.
- Galpern, Steven G. *Money, Oil, and Empire in the Middle East: Sterling and Postwar Imperialism, 1944-1971*. Cambridge: Cambridge University Press, 2009.
- Heiss, Mary Ann. *Empire and Nationhood: The United States, Great Britain, and Iranian Oil, 1950-1954*. New York: Columbia University Press, 1997.
- Inboden, William. *Religion and American Foreign Policy, 1945-1960: The Soul of Containment*. Cambridge University Press, 2010.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. Urbana and Chicago: University of Illinois Press, 2013.

- Keddie, Nikki R, and Yann Richard. *Modern Iran: Roots and Results of Revolution*. New Haven: Yale University Press, 2006.
- Moretti, Franco. *Distant Reading*. New York: Verso, 2013.
- Office of the Historian, Bureau of Public Affairs. "About Us." Office of the Historian, 2019. <https://history.state.gov/about>.
- Painter, David S. *Oil and the American Century: The Political Economy of U.S. Foreign Oil Policy, 1941-1954*. Baltimore: Johns Hopkins University Press, 1986.
- Preston, Andrew. *Sword of the Spirit, Shield of Faith: Religion in American War and Diplomacy*. New York: Knopf Doubleday Publishing Group, 2012.
- Westad, Odd Arne. *The Cold War: A World History*. New York: Basic Books, 2017.
- Yaqub, Salim. *Containing Arab Nationalism: The Eisenhower Doctrine and the Middle East*. Chapel Hill: University of North Carolina Press, 2004.